



Indian Government Interference in Internet Tamil

**(Tamil Language in the Age
of Computers, Electronics
and Internet)**

Thanjai Nalankilli

Indian Government Interference in Internet Tamil (Tamil Language in the Age of Computers, Electronics and Internet)

by
Thanjai Nalankilli

Copyright Thanjai Nalankilli 2019

This book or any chapter in the book may be copied, distributed, reposted, reprinted, translated or shared in print, electronic, digital, internet or other media. No permission needed from copyright holder.

Table of Contents

[Preface](#)

[1. Indian Government, Tamil Unicode and Devanagari Script \(How the Indian Government "Down-Graded" Tamil Unicode?\) \(by Thanjai Nalankilli\)](#)

[2. Indian Government Attempts to "Pollute" Tamil Unicode with Grantha Characters \(by Thanjai Nalankilli\)](#)

[3. Indian Government Pollutes Tamil Website with Hindi-Sanskrit Words \(by Thanjai Nalankilli\)](#)

[4. Hindi-centric Automated Computer Translation of Indian Languages \(Tamil, Kannada, Malayalam, Telugu ...\) \(by Thanjai Nalankilli\)](#)

[List of More free E-Books from Us](#)

[Back to Table of Contents](#) or Scroll down for the Preface

Preface

This is the digital age. This is the age of electronics and Internet. Languages should adapt to electronic media in order to survive and thrive in the decades to come. Tamil is the first South Asian language to enter the Internet/web. Neither the Tamil Nadu State government nor Indian government is responsible for this. Credit should go to Nanyang Technological University and National University of Singapore. We are grateful to the Government of Singapore for funding the projects.

Large amount of Tamil content is available to day on the Internet. Today we can use Tamil in most major international websites; you can post your views in Tamil; publish your books in Tamil; you can read, buy and sell things in Tamil. There is one family of websites where use of Tamil is limited or simply not available. These are Indian government websites. Out of the hundreds of Indian government websites, you will find Tamil content or use Tamil in only a handful of them. You can find a number of examples of the discrimination of non-Hindi languages in Reference 1. This book is not about lack of Tamil on Indian government websites.

Indian government is making every effort to degrade and denigrate Tamil on the Internet. Since the largest population of Tamils live in India, Internet standards setting organizations such as International Organization for Standardization come to Indian government for recommendations on standards for using Tamil on the Internet. Indian government cunningly gives recommendations that degrade and denigrate Tamil and show Tamil script as a subset to Devanagari script used by Hindi and Sanskrit. Tamil is an orphan language in the international arena without a country to safeguard it. It is a motherless child. Indian government treats Hindi and Sanskrit as her daughters and all other languages spoken in India as stepdaughters. What can we do to protect the purity and dignity of Tamil?

Reference

1. "Hindi Hegemony on Indian Government Websites (Chapter 11)", Hindui Imposition Papers (Volume 15), Free e-book available where you downloaded this book.

[Back to Table of Contents](#) or Scroll down for the next chapter

1. Indian Government, Tamil Unicode and Devanagari Script

(How Indian Government "Down-Graded" Tamil Unicode?)

Thanjai Nalankilli

[First Published: January 2017]

OUTLINE

Executive Summary

1. What is Unicode?
2. Indian Government Makes Tamil Script a Subset of Devanagari
3. How Tamil Unicode was "Degraded"?
4. An Un-needed Crutch for Tamil
5. Indian Government Promoting Devanagari Script
 - 5.1. The Case of Konkani Language
 - 5.2. The Case of Kashmiri Language
 - 5.3. The Case of Indian Posts and Telegraph
6. A Request to Computer Standards Organizations

ABBREVIATION

ISO - International Standardization Organization (International Organization for Standardization)

EXECUTIVE SUMMARY

Tamil script is better suited for computer rendering than Devanagari script because Tamil does not have conjunct consonants. Because of Indian government's incorrect recommendation that Tamil script be treated as a subset of Devanagari script, Tamil is brought down a notch to the same level as Devanagari. Intentional or not, it is unfortunate. We request that computer standards organizations consult the state governments (organized on linguistic basis) on matters relating to Indian languages.

1. What is Unicode?

Unicode is a standard system adopted by major computer software developers to create documents in different languages using computers. For those who want a more technical definition: Unicode is an international encoding standard to support the processing and display of written texts of diverse languages across different platforms and programs.

2. Indian Government Makes Tamil Script a Subset of Devanagari

The Unicode consortium approached Indian government for its recommendations for the various "Indian languages". (We want to point that there is no linguistic family called

"Indian languages". The phrase "Indian languages" refers to the languages spoken within the country of India that was formed in 1947 following the end of British rule over South Asia. There are many languages in this country, some dating back to thousands of years and some hundreds of years. There is no common script for all these languages; neither do all the scripts fall under some "script family".)

Indian government asked the Unicode consortium to treat all Indian language scripts as subsets of Devanagari script although linguistically Devanagari is not a superset nor super-script for Tamil script (and may be some other Indian languages too).

3. How Tamil Unicode was "Degraded"?

Because the Indian government asked the Unicode consortium to treat all Indian language scripts as subsets of Devanagari script although linguistically (scientifically) Devanagari is not a superset nor super-script for Tamil script, writing and displaying Tamil script on computers less efficient and less optimal (sub-optimal). [Author borrowed the terms superscript and sub-optimal from Mani Manivannan's posts in Facebook in October 2014.] The following paragraph is based on Mr. Manivannan's 2014 interview to a Tamil magazine [Reference 1] and his Facebook posts in October 2014.

"A major difference between Tamil and Devanagari scripts is that Tamil script does not have conjunct consonants like Devanagari. Because Unicode Tamil script does not have conjunct consonants, Unicode could have used a linear arrangement of Tamil consonants. This would have made it unnecessary for the use of "complex rendering engine" (Devanagari requires it. If properly done Tamil would not have.) But because the Indian government asked Unicode consortium to make Tamil script a subset of Devanagari script, Tamil also now requires complex rendering engine. This makes display of Tamil text in some word processing software incorrect and searching Tamil text in Adobe PDF software Acrobat difficult, until these software implement complex script support." A more detailed discussion may be found in Reference 1; we suggest that those who know Tamil read that reference for more information. [Mani M. Manivannan was chairman of Tamil Unicode Working Group of the International Forum for Information Technology in Tamil (INFITT)]

4. An Un-needed Crutch for Tamil

By asking the Unicode to treat Tamil script as a subset of Devanagari script, Indian government made Tamil wear a crutch it does not need. Let me explain. Tamil script is better suited for computers than Devanagari because it does not have conjunct consonants. Scripts that have conjunct consonants (for example, Devanagari) need complex rendering engine (an additional software) to display the letters properly on screen and in searching text. By making Tamil script a subset of Devanagari, Indian government had made it necessary for Tamil also to require complex rendering engine (a necessary "crutch" for Devanagari but an unnecessary crutch for Tamil).

5. Indian government promoting Devanagari Script

Indian government asking the consortium to treat Devanagari script as the superscript for all Indian languages seems to be an intentional effort to set Devanagari script above all else. Devanagari script is used to write Sanskrit and Hindi, the two languages favoured by Indian government. Indian government pushing Devanagari script over and above other Indian language scripts is not new.

5.1. The Case of Konkani Language

Konkani language is spoken mostly in southern India. Majority of Konkani speakers live in Karnataka State (Kannada is the primary state language). Konkani does not have its own script. Most Konkani literature is written in Kannada script, with some literature in Malayalam, Roman (English), Arabic and Devanagari. In 2016, Indian Government's Sahitya Academy made it mandatory for Konkani submissions for the prestigious Sahitya Academy Awards be in the Devanagari script (Bangalore Mirror; May 2, 2016). Sahitya Academy's order is like imposing Devanagari script on the Konkani language. Some Konkani scholars sued the academy and it is still pending in court.

5.2. The Case of Kashmiri Language

Indian government proposed in 2005 that Devanagari be used as an alternative script for Kashmiri language although there is a distinct Kashmiri script for over five centuries. The proposal was dropped because of opposition among Kashmiri speakers. Indian government again brought out the same proposal in 2016. Kashmiri writers and poets opposed the idea again (Kashmir Reader; May 24, 2016).

5.3. The Case of Indian Posts and Telegraph

When telegraph was a popular form of long-distance communication (before the days of fax and e-mail), Indian government allowed only English and Hindi. When other language groups wanted telegraph in their languages too, Indian government told them that they may telegraph in their language if they write the message in Devanagari script; another attempt to force people to use Devanagari script. Indian government yielded after much protests from Tamil Nadu and Tamil was also made available in some telegraph offices.

Thus asking Unicode consortium to use Indian language scripts as subsets of Devanagari seems to be yet another attempt to thrust this script into other languages.

6. A Request to Computer Standards Organizations

Indian government is not the guardian of all Indian languages. Guardians of Indian languages are the state governments where the respective languages are spoken. Karnataka for Kannada, Kerala for Malayalam, Maharashtra for Marathi, Tamil Nadu for Tamil, Andhra and Telengana for Telugu, etc. So international standardization organizations like ISO should consult state governments (if necessary, through the Indian

government) and take their recommendations into account. This is the only way to protect the uniqueness and integrity of Indian languages.

Reference

1. <http://www.yarl.com/forum3/index.php?showtopic=144938>

[Back to Table of Contents](#) or Scroll down for the next chapter

2. Indian Government Attempts to "Pollute" Tamil Unicode with Grantha Characters

Thanjai Nalankilli

[First Published: December 2018]

OUTLINE

1. Introduction
2. Background
3. Attempted Pollution in 2010
4. A Second Attempt to Mix Grantha and Tamil Characters
5. Indian Government and Tamil Nadu State Government Enter the Debate
6. A Sinister Attempt to Degrade Tamil and Elevate Sanskrit?

1. Introduction

Unicode is a standard system adopted by major computer software developers to create documents in different languages using computers. This system is managed and maintained by the Unicode Consortium. Like most other languages, Tamil also has a block allotted to it in the Unicode system. Tamil Nadu State Government is a voting member of the Unicode Consortium since May 2007. Indian Government is a voting member even before that.

2. Background

Unicode system makes it possible to type the 247 Tamil letters (characters) in computers. In addition Unicode also includes the five Grantha letters (characters) "ja, sha, sa, ha and sri" used in some commonly used Sanskrit-origin words or Sanskritized Tamil words such as "rajan" or "kashtam"; Tamil Nadu State Government has accepted this inclusion. I want to point out that these five are not Tamil letters (characters) but Grantha letters used to spell some Sanskrit-origin words or Sanskritized Tamil words. Tamil has only 247 letters and these 5 are not among them.

This article is not about the inclusion of these 5 Grantha letters in the Tamil Unicode block. This article about an attempt by some private individual(s) and Indian government to add 26 more Grantha letters, which not even one in a million Tamil people have heard about. It was used only twice in two obscure books in the past 100 years. (1 million = 10 lakhs).

3. Attempted Pollution in 2010

A gentleman named Sri Ramana Sharma proposed to the Unicode consortium in 2010 that space be allocated in the Tamil Unicode block for 26 Grantha characters. Many

Tamil scholars believe that it is an unnecessary addition because these letters are not used by Tamil writers. Proponents of adding these 26 characters could show only two examples of their usage anywhere in the last 100 years. The two books are (1) Bhoja Sharitham by T.S. Narayan Shastry (1916), and (2) Shiva Manasa Puja Kirthanas mattrum Aathma Vidya Vilasa by Sri Sadasiva Paramendra (1951).

Many Tamil scholars, organizations and Tamil Nadu state government opposed the addition. Professor G. Hart of University of California (Berkeley, United States of America (USA)) who has extensive knowledge of Sanskrit and Tamil opposed the proposal and wrote [Reference 1], "Mr. Sri Rama Sharma proposed to the Unicode consortium that space be allocated in the Tamil Unicode Block for Grantha (Sanskrit) characters...It would, in my view, be a serious mistake to include Sanskrit sounds (except for those in general use, like "ja") in Tamil unicode... Keeping Grantha and Tamil separate, with separate Unicode blocks, should satisfy everyone. If one looks through Sangam literature or Kampan, there is not a Grantha letter to be found. In modern Tamil books, the only Grantha letters are those few that are needed for foreign words [The professor is referring to the 5 Grantha letters already included in Tamil Unicode block]. There is absolutely no need to expand the Tamil Unicode slots to include the unused Grantha letters [He is referring to the 26 Grantha letters]. The inclusion of Sanskrit sounds in Tamil, where necessary, can easily be accomplished by combining Tamil and Grantha, leaving Tamil as it is at present."

Because of the opposition from Tamil scholars, organizations and Tamil Nadu State Government, the proposal was not accepted.

4. A Second Attempt to Mix Grantha and Tamil Characters

The second attempt was to create a Grantha block in Unicode. Tamil scholars or Tamil Nadu state government have no objection to it. What we object is the proposal to include 5 Tamil characters (letters) "a, o, za, Ra, na" into the Grantha Unicode block. If they cannot mix Tamil and Grantha characters by adding 26 Grantha characters into Tamil Unicode block, they would mix Tamil and Grantha characters by adding 5 Tamil characters (letters) into the Grantha block. Tamil scholars objected to this also.

5. Indian Government and Tamil Nadu State Government Enter the Debate

Although the proposal to include 5 Tamil letters into the Gratha block came from private individuals, Indian government supported that proposal. Indian government wrote to Unicode on September 6, 2010 supporting the inclusion of the 5 Tamil letters into Grantha block. Indian government did not consult Tamil scholars or even inform Tamil Nadu government of their support for mixing Tamil characters with Grantha. When Tamil scholars came to know of Indian government support, they opposed it and also informed the Tamil Nadu government.

The then Tamil Nadu Chief Minister M. Karunanidhi wrote to Indian Government Minister for Communications and Information Technology on November 6, 2010

[Reference 2]. He wrote, "... This proposal has raised considerable concern from a wide cross-section of Tamil community from around the world. They have indicated that sufficient consultations have not taken place with eminent Tamil language scholars, before submitting the proposal. In particular, considerable reservations have been expressed about inclusion of five Tamil characters into the Grantha code places."

6. A Sinister Attempt to Degrade Tamil and Elevate Sanskrit?

Several Tamil Unicode experts have pointed out that there is no technical advantage in mixing Tamil and Grantha characters rather than have two separate unicode blocks with no overlap or duplication. For example, Director of Tamil Virtual Academy Dr. P.R. Nakkeeran wrote [Reference 3], " Adding characters that are native to Tamil script but not part of the Grantha script can potentially lead to confusion when digitizing ancient Tamil inscriptions that have Grantha characters". Readers interested in the technical details may read References 3, 4 and some of the references listed in Reference 4.

There are Sanskrit enthusiasts in India who want to elevate Sanskrit as the supreme language of India. Tamil writers had used Sanskrit words in their writings and Sanskrit writers had use Tamil words in their wrings. This does not mean one is the mother of the other or one is older or superior to the other. Many people use English words today while speaking/writing Tamil, Telugu, Hindi, etc. Does this mean English is the mother of all these languages? Absolutely, not.

Mother-daughter relationship has to be established on the basis of etymology, grammar and other factors. I am not a linguist and I cannot elaborate further on it. So I quote Dr. G. Hart (Professor of Tamil at University of California, Berkeley, USA). Professor Hart is uniquely qualified to comment on the relationship between Sanskrit and Tamil. He knows Sanskrit and Tamil well. He got his doctorate degree in Sanskrit from one of the top universities in the world - Harvard University, USA. He has also studied Latin, Greek and Tamil. He has translated to English some classic Tamil literature. He published books on Tamil and Sanskrit. He served as Professor of Sanskrit at University of Wisconsin (USA) before becoming Professor of Tamil at University of California. Here is what he has to say [Reference 5]: "Tamil had the good fortune to gain an extensive written literature before the Sanskrit juggernaut became irresistible. Its early works owe virtually nothing to Sanskrit. ... Because Tamil developed its own identity so early, it remained relatively immune to the influence of Sanskrit. It retained (and retains) its own writing system that genuinely fits the pronunciation of the language ... The early origins of Tamil and of its writing system have helped it keep its separate identity from Sanskrit.... Its separate identity and character have been cultivated and preserved from its beginnings to the present".

This is the view of a scholar in both Sanskrit and Tamil. Yet some Sanskrit enthusiasts, including some Indian government ministers, continue to assert Sanskrit is the supreme language of India, and all other languages owe to Sanskrit.

A former Indian minister for human resources development (HRD), Mr. Murli Manohar Joshi, said, "Every Indian language is associated with Sanskrit. Some are directly derived from it and some have a large component as their diction, vocabulary and grammar" (Rediff on the Net; February 23, 2002). India's home minister Rajnath Singh said in September 2015, "mother language of all Indian languages Sanskrit" (Hindustan Times; September 16, 2015). There is no basis for these statements. Yet some Sanskrit enthusiasts continue to repeat it; unfortunately some of these people are in positions of power to set language policies in India. The fact that all three ministers we quoted are Bharatiya Janata Party (BJP) leaders does not mean that Indian government supports Sanskrit supremacy policies only when BJP is in power. In fact Indian government sent Unicode consortium the proposal to include 5 Tamil characters in the Grantha Unicode block (discussed in Section 4) when Congress Party was in power. One party may be more open and vocal but both parties have more or less the same language policy.

Attempts to include some Grantha characters in Tamil Unicode block and some Tamil characters in the Grantha Unicode block seem like an attempt to mix Tamil characters with Grantha and then claim years later that Tamil is not self-sufficient unlike Sanskrit and thus assert falsely Sanskrit supremacy. So we shall continue to oppose mixing Tamil and Grantha characters in Unicode by the inclusion of 26 Grantha characters into Tamil Unicode block or including 5 Tamil characters into the Grantha Unicode block.

REFERENCES

1. G. Hart, Grantha letters in Tamil, October 29, 2010
[<https://web.archive.org/web/20120312064930/http://tamil.berkeley.edu/grantha-letters-in-tamil>]
2. Tamil Nadu Chief Minister M. Karunanidhi's letter to Indian Government Minister for Communications and Information Technology dated November 6, 2010
[<https://web.archive.org/web/20170304040656/https://www.unicode.org/L2/L2010/10464-tamil-nadu.pdf>]
3. Dr. P.R. Nakkeeran's Letter to Dr. Swaran Latha dated November 1, 2010 (L2/10-457)
[<https://web.archive.org/web/20170304042357/http://www.unicode.org/L2/L2010/10457-grantha-fdbk.html>]
4. Dr. B. Eraiyyarasan's comments on Tamil Unicode and Grantha proposals (L2-11/055)
[<https://web.archive.org/web/20170304042602/https://unicode.org/L2/L2011/11055-tamil-grantha.pdf>]
5. G. Hart, Sanskrit and Tamil, November 25, 2010
[<https://web.archive.org/web/20130306110903/http://tamil.berkeley.edu/sanskrit-and-tamil>]

[Back to Table of Contents](#) or Scroll down for the next chapter

3.

Indian Government Pollutes Tamil Website with Hindi-Sanskrit Words

Thanjai Nalankilli

[First Published: February 2019]

OUTLINE

1. Introduction
2. Details
3. Not-so-hidden Agenda of the Indian Government: Degradation of Tamil and Elevation of Hindi-Sanskrit
4. What Shall We Do?

1. Introduction

The Central Institute of Indian Languages (CIIL) in Mysuru, Karnataka State funded by the Indian Government Ministry of Human Resource Development (HRD) has a project "Bharatavani: Knowledge through Indian languages". Its website has dictionaries, learning tools, books, etc. in various Indian languages. It is a good effort but Indian government imposes Hindi even in this worthwhile project; it unnecessarily, purposively and sinisterly degrades Tamil by mixing Hindi-Sanskrit words in the Tamil web page. Here are the details.

2. Details

The following information is based on what the author saw at their website on February 22, 2018. If you go to the main page [Reference 1] you will see Hindi and English but there are also links to various Indian language pages. That is good. Problem is in the Tamil web page. In the Tamil opening page [Reference 2], there are five main links. The problem is they are Hindi words written in Tamil script. These are some of the links in the Tamil page as they appear in that page [You notice that the links are in Tamil script (letters) with English script in parentheses]. I have presented it in both graphic format (jpg format) and in Tamil Unicode script because some e-readers do not show graphics imbedded in text and some e-readers do not show Tamil Unicode letters at all or show as some garbled script. Some e-readers show them both correctly.

In graphic format (jpg file):

1. பாஷா கோஷா
(Bhashakosha)
2. பாத்யபுஸ்தக கோஷா
(Textbooks)
3. ஞான கோஷா
(Jhanakosha)
4. பஹுமாத்யம கோஷா
(Multimedia)
5. ஷப்த கோஷா
(Dictionary)

In Tamil Unicode letters:

1. பாஷா கோஷா
(Bhashakosha)
2. பாத்யபுஸ்தக கோஷா
(Textbooks)
3. ஞான கோஷா
(Jhanakosha)
4. பஹுமாத்யம கோஷா
(Multimedia)
5. ஷப்த கோஷா
(Dictionary)

No Tamil will understand what these links mean unless he/she knows Hindi. Three of the five links are translated to English but remaining two links are Hindi-Sanskrit words written in English script. So people who know Tamil only would not understand any of the links. People who know English would understand 3 links. Only people who know Hindi would understand all 5 links.

For example, the second link says, "பாத்யபுஸ்தக கோஷா (Textbooks)"

"பாத்யபுஸ்தக கோஷா (Textbooks)". I understand it because I know English.

What about Tamils who do not know English? Tamil translation for textbooks is பாட

நூல்கள் (paada noolkaL) பாட நூல்கள் (paada noolkaL). Any Tamil with a fifth

grade education will understand it. No Tamil without a knowledge of Hindi or Sanskrit

will understand பாத்யபுஸ்தக கோஷா பாத்யபுஸ்தக கோஷா .

This is not accidental or incidental that Hindi-Sanskrit words are unnecessarily infused into Tamil.

3. Not-so-hidden Agenda of the Indian Government: Degradation of Tamil and Elevation of Hindi-Sanskrit

There are Sanskrit enthusiasts in India who want to elevate Sanskrit as the supreme language of India. Some of them are powerful politicians including Indian government ministers. One thing that stands in the way of declaring Sanskrit supremacy is Tamil that is old, still alive and independent of Sanskrit or other languages. Professor G. Hart is a scholar in both Sanskrit and Tamil, has taught Sanskrit and Tamil at American universities and has published books on both languages. He said the following: "Tamil had the good fortune to gain an extensive written literature before the Sanskrit juggernaut became irresistible. Its early works owe virtually nothing to Sanskrit. ... Because Tamil developed its own identity so early, it remained relatively immune to the influence of Sanskrit. It retained (and retains) its own writing system that genuinely fits the pronunciation of the language ... The early origins of Tamil and of its writing system have helped it keep its separate identity from Sanskrit.... Its separate identity and character have been cultivated and preserved from its beginnings to the present" [Reference 3].

This is the view of a scholar in both Sanskrit and Tamil. Yet some Sanskrit enthusiasts, including some Indian government ministers, continue to assert Sanskrit is the supreme language of India, and all other languages owe to Sanskrit.

A former Indian minister for human resources development (HRD), Mr. Murli Manohar Joshi, said, "Every Indian language is associated with Sanskrit. Some are directly derived from it and some have a large component as their diction, vocabulary and grammar" (Rediff on the Net; February 23, 2002). India's home minister Rajnath Singh said in September 2015, "mother language of all Indian languages Sanskrit" (Hindustan Times; September 16, 2015). There is no basis for these statements. Year from now people like these could show the web site to declare Tamil needs to use these Hindi-Sanskrit words and cannot survive without them in the modern computer age.

4. What Shall We Do?

An article like this does not get wide circulation. We ask Tamil people and Tamil activist groups to write/contact Tamil Nadu State government and pro-Tamil politicians about this Central Institute of Indian Languages (CIIL) website. Tamil Nadu state government and pro-Tamil politicians should publicly criticize it and write to Indian government and the Central Institute of Indian Languages (CIIL) to use Tamil words instead of the Hindi-Sanskrit words throughout the website. When we say Tamil words, we mean pure Tamil words and not Sanskritized or Englified Tamil words. If necessary, they should consult or hire Tamil scholars to create Tamil pages on their websites.

REFERENCES

1. Bharatavani Main Entry Page

[<https://web.archive.org/web/20180222133547/http://bharatavani.in:80/>]

2. Bharatavani Tamil Entry Page

[<https://web.archive.org/web/20180108024441/http://tamil.bharatavani.in:80/#books>]

3. G. Hart, Sanskrit and Tamil, November 25, 2010

[<https://web.archive.org/web/20130306110903/http://tamil.berkeley.edu/sanskrit-and-tamil>]

[Back to Table of Contents](#) or Scroll down for the next chapter

4. **Hindi-centric Automated Computer Translation of Indian Languages (Tamil, Kannada, Malayalam, Telugu ...)**

Thanjai Nalankilli

[First Published: November 2016]

OUTLINE

Abbreviations

Executive Summary

A Note

1. Introduction
2. A Few Questions to TIFAC
3. There is no Language Family Called "Indian Languages" (from Linguistic Perspective)
4. English-to-Telugu versus English-to-Hindi-to-Telugu
5. Quality of Computer Translations Degraded by this Multi-Step Approach
6. Oppose the Meta-Language Approach

ABBREVIATIONS

ISO - International Standardization Organization (International Organization for Standardization)

TIFAC - Technology Information, Forecasting and Assessment Council

A NOTE

Author of this article is from Tamil Nadu and so Tamil language is used as an example but the conclusions of this article are applicable to some other Indian languages also.

EXECUTIVE SUMMARY

Executive Director of an Indian government affiliated organization, TIFAC, had suggested that automated computer translation of all Indian languages from English go through a meta-language. Although the meta-language was not specified, based on our experience with the Indian government, we have reason to expect that meta-language would be either Hindi or Sanskrit. This Hindi-centric approach should be opposed because all Indian languages would forever become dependent on Hindi and the quality of translations would also suffer.

1. Introduction

Indian language localization community (those involved in creating Internet content in local languages like Hindi, Telugu, Tamil, Marathi, Kannada, Bengali ...) met in New

Delhi (India) on September 24-25, 2016. A proposal by the executive director of TIFAC, Dr. Prabhat Ranjan, raised concern in this writer. "Ranjan's team found English to Hindi translation easier when documents are first translated into another Indian language. Based on this experience, Ranjan bounced the idea of agreeing on a meta language to ease the translation process." [Reference 3]. Reference 3 did not mention what was the "another Indian language" that Dr. Ranjan's team used in the English to Hindi translation. Also, Dr. Ranjan did not suggest a choice for the meta-language.

Our fear is that the Indian government or some other organization or individual may try to elevate Hindi or Sanskrit as the meta language on the basis of Dr. Ranjan's statement, although he did not suggest a choice for the meta-language at the meeting.

TIFAC is an autonomous body under Department of Science and Technology of Government of India. Indian government efforts to establish Hindi/Sanskrit as a super-language over all other Indian languages is no secret. Indian government Home Minister Rajnath Singh said that Sanskrit is the mother of all Indian languages and he considers Hindi as the elder sister of all regional languages because it is closer to Sanskrit (Hindustan Times; September 16, 2015). This statement may be true for SOME Indian languages but false for others, for example, Tamil. One should not bunch linguistically unrelated languages into a single family.

This effort to establish Hindi/Sanskrit as a meta language through which all automated computer translations flow should be opposed, and international standardization bodies like ISO should not accept it. These bodies should not become unwitting tools in the hands of the Indian government or other vested interests.

2. A Few Questions to TIFAC

Summary of Dr. Prabhat Ranjan's speech says, "Our team found English to Hindi translation easier when documents are first translated into another Indian language". What was that "another Indian language"? Was it Sanskrit or some other Indo-Aryan language? It would not be a surprise if that "another Indian language" was Sanskrit or another Indo-Aryan language. Not all Indian languages are related to Sanskrit/Hindi. His conclusion would not hold for those languages. Tamil, for one, has very little in common with Sanskrit/Hindi. Professor George L. Hart of University of California, Berkeley, United States of America (USA) said, "Tamil is extremely old (as old as Latin and older than Arabic); it arose as an entirely independent tradition, with almost no influence from Sanskrit or other languages" [Reference 1]. He says, "The early origins of Tamil and of its writing system have helped it keep its separate identity from Sanskrit.Its separate identity and character have been cultivated and preserved from its beginnings to the present, and they will be preserved." [Reference 2].

Just because a small percentage words are common between Sanskrit and Tamil, neither may be concluded as derived from the other. There are so many English words used in Tamil these days; it does not mean Tamil is derived from English.

3. There is no Language Family Called "Indian Languages" (from Linguistic Perspective)

There is no language family or language group called "Indian Languages" from the perspective of linguistics. Indian languages is a political marker or geographic marker. The two main language families in India (or South Asia) are Indo-Aryan and Dravidian. Indo-Aryan languages are spoken by about 78.05% of Indians and Dravidian languages spoken by about 19.64%. The remaining 2.31% people speak languages belonging to Austro-Asiatic, Sino-Tibetan, Tai-Kadai and a few other minor language families [Reference 4]. Hindi and Sanskrit and a number of other languages belong to the Indo-Aryan family; Tamil, Telugu, Kannada, Malayalam and some other languages belong to the Dravidian family.

Any suggestions or recommendations of translating English to any Indian language through Sanskrit/Hindi (or any "Indian meta language") is without merit. It has no scientific rationale. Using data, if any exists, of translating English to an Indo-Aryan language through Hindi/Sanskrit to justify translating English to a Dravidian language through Hindi/Sanskrit would be voodoo science, voodoo-linguistics. It is unacceptable.

4. English-to-Tamil versus English-to-Hindi-to-Tamil

For the sake of argument let us say than Hindi-to-Tamil automated computer translation is cheaper and more accurate than English-to-Tamil translation. Is English-to-Hindi-to-Tamil translation cheaper and more accurate than English-to-Tamil? I doubt it. Such a scenario should be studied and established for each and every Indian language--be it Telugu or Malayalam or Kannada or Telugu or Bengali or Oriya or Manipuri. We cannot accept this meta language suggestion based on undemonstrated hypotheses.

5. Quality of Computer Translations Degraded by this Multi-Step Approach

It is an established fact that quality is degraded during automated computer translations. If one translates a text from English to Russian and then translate back the translated Russian text to English, it would not be the same as the original English text. The same would happen in the proposed meta language three-step approach. A direct translation would be truer to the original text than an indirect three-step approach suggested by TIFAC executive director Dr. Ranjan. That is, a Tamil translation directly from English would be closer to the original English text than first translating English to Hindi and then Hindi to Tamil. So Hindi would end up with higher quality translations and the other Indian languages like Marathi, Oriya, Bengali, Kannada, Malayalam, Telugu would end up with lesser quality translations. We cannot allow such a systematic two-tier approach, a higher level of translation for Hindi/Sanskrit and a lower level of translation for the other languages.

6. Oppose the Meta-Language Approach

We explained in the preceding sections how centralizing translations through Sanskrit/Hindi would be detrimental to many Indian languages. Politicians, scholars and the public should oppose this approach. Tamil scholars and computer specialists active in automated computer translation of Indian languages should contact the Unicode standards organizations and request them to contact not only the Indian government but also state governments on language related matters. India is a multi-lingual country. States were reorganized on the basis of languages in the 1950s so that the major languages have a state where each major language can flourish. State governments are there to nurture and protect them. So it is appropriate that recommendations on languages come from the states.

REFERENCES

1. Statement on the Status of Tamil as a Classical Language (by George L. Hart), University of California at Berkeley, April 2000.
<https://web.archive.org/web/20100308114310/http://tamil.berkeley.edu/Tamil%20Chair/TamilClassicalLanguage/TamilClassicalLgeLtr.html>
2. Sanskrit and Tamil (by George L. Hart), University of California at Berkeley, November 25, 2010.
<https://web.archive.org/web/20130306110903/http://tamil.berkeley.edu/sanskrit-and-tamil>
3. GILT-Conference
<https://web.archive.org/web/20161004033502/https://opensource.com/life/16/10/gilt-conference>
4. Languages of India
https://web.archive.org/web/20190112112611/https://en.wikipedia.org/wiki/Languages_of_India

[Back to Table of Contents](#) or Scroll down for "List of More free E-Books from Us"

More Free E-books

The following e-books are available FREE on the Internet in the same web site you found this book. If not, you may also find it at **<https://archive.org/details/texts>** Search by book title.

Also, if you like this book please tell your friends or e-mail them the book or link to the book. You may also post links and comments about the book on media sites (facebook, twitter, ...) or discussion forums of your choice. Thank you.

Tamil Language

Tamil Language: River Valley Civilization to Silicon Valley Civilization

Indian Government and Tamil Language (Ancient Rivalries and Current Tug of War)

Tamil in Tamil Nadu Schools, Colleges and Universities

Tamil Today: Some Thoughts and Musings

Indian Government Interference in Internet Tamil: (Tamil Language in the Age of Computers, Electronics and Internet)

India - Current Affairs

The Failed Indian Democracy: Devolution is the Solution

Economic Discrimination of South India (With Examples from Tamil Nadu)

Hindi Imposition Papers (15 volumes)

Hindi, India and the United Nations: Opposing View from Non-Hindi Indians

Battle for Tamil Eelam

Tamil Nadu and the Battle for Tamil Eelam (Sri Lanka)

India and the Battle for Tamil Eelam (Sri Lanka)

United States of America and the Battle for Tamil Eelam (Sri Lanka)

International Community and the Battle for Tamil Eelam (Sri Lanka)

Thoughts and Musings on the Battle for Tamil Eelam (Sri Lanka)

Tamil Nadu

Early History of Dravidian Parties in Tamil Nadu: 1916-1959

A Political History of Dravidian Movement and Parties in Tamil Nadu

Economic Discrimination of South India (With Examples from Tamil Nadu)

Indian Interference in the Internal Affairs of Tamil Nadu

Tamil Nadu: Historical Perspectives, Distortions and Blackouts

A Political Biography of Tamil Nationalist Poet Pavalareru Perunjchiththirananar

Dark Clouds over Tamil Nadu (Cauvery Water, NEET, Jallikattu, Navodaya Schools and Fishermen Shootings)

The Two Golden Ages of Tamil Nadu and the Current Dark Days

Lighter Side of Life and Politics in Tamil Nadu

Books in Tamil

தமிழ்நாட்டின் இரு பொற்காலங்களும், இன்றைய இருள் காலமும் (வரலாறு) - தஞ்சை நலங்கிள்ளி

The Two Golden Ages of Tamil Nadu and the Current Dark Days (History) (Tamil book) by Thanjai Nalankilli

தமிழ்நாடு இந்தி எதிர்ப்புப் போராட்டங்களும், தீக்குளிப்புகளும் - தஞ்சை நலங்கிள்ளி
Tamil Nadu Anti-Hindi Agitations and Self-Immolations (Tamil book) by Thanjai Nalankilli

END OF BOOK

[Back to Table of Contents](#)